NEANIAS Space-Planets Pipelines

Carlos H. Brandt, Angelo P. Rossi, Laura Vettorello, Simone Mantovani

NEANIAS Space/Planets working group is working to provide high-level services to otherwise technically bound Planetary image data processing algorithms; the goal is to offer science-level data and common but computer expensive tasks through an intuitive and interactive user interface so a broad community of data scientists, engineers and enthusiasts can benefit from some of the best observational space mission humankind has deployed on Mars and the Moon in the last decades. In this article we present the steps necessary to produce the image products already available through our services, as well as the interactive actions users can launch for generating custom products.

Introduction

One of the guidelines of NEANIAS services is to provide FAIR (Findable, Accessible, Interoperable, Reproducible) data. Many are the aspects concerning those properties of a data set, and the discussions addressing them branch different implementations depending on the field of science. Regardless, though, openness and high-level data are key to address the ultimate goal of a data set: to be used as widely as possible so that data can add as much value as possible through the best of its analysis and derived knowledge. Openness is fundamental for it allows unrestricted access and so is fundamental for any unbiased use we may want to make out of it.

The concept of "high-level" of a data product is somewhat abstract – and, again, field-dependent – but we can simplify it for "easy to use". A JPEG image is an example of a high-level format for representing an image since it has broad support for the visualization of its content in our digital devices. For geographical images such of a *high-level* format is GeoTIFF, an extension of the well-known TIFF format with geo-location embedded. GeoTIFF can be visualized by virtually any image viewer, capable of rendering TIFF files, and the specific manipulation of the geographical information in it (e.g, re-projection to different cartographic representations) can also be done by many available tools.

For different reasons, though, scientific data archives do not usually provide their data in some high-level data formats but rather data formats that can be handled by a few specific software tools. This is the case for datasets provided by the institutions in front of NASA and ESA missions providing, for instance, Mars and Moon images. One of the reason for the use of unpopular data formats in scientific data archives regards the complexity and volume of information stored in those data sets: as a source of truth, these so called (scientific) *source* data archives have the duty of providing every and all piece of information captured by the science mission behind it, rendering the data and metadata justly adequate for niche-specific demands.

Here is where NEANIAS comes in: to fill the gap between low-level, niche-specific data products and data processing methods capable of being provided by high-level representations through it services. In particular, the Space-Planets service developed by the

Jacobs University Bremen[1] and MEEO [2]remove all the boilerplate necessary for elevating the level of the main Mars and Moon surface observation imagery data provided by NASA and ESA academical institutions to the non-expert user and provide common data analysis tools as a service through the ADAM[3] platform.

Data processing pipelines

To provide the data to the user through the ADAM platform, the images available at the source data archives must go through a series of processing steps to reach the high-level state suitable for our users. Such processing steps are known as processing pipelines in analogy to a stream of water flowing through pipes across a series of nodes.

High-level data is also referred to as *reduced* data products, which are images with their signal corrected by instrumental effects, such as instrumental noise, and calibrated to provide homogeneous physical quantities such as the surfaces reflectance given that the observation itself is influenced by atmospheric effects of the incidence of Sun light to some degree. Reduced data products, as provided by us, besides having their signal calibrated to account for empirical observations effects are also projected from the instrument frame of reference – which does not about cartographic projections - to properly represent the data in some known cartographic projection like equirectangular or sinusoidal and allow such data to be handled by cartographic-aware interface, such as ADAM-Space[4].

The processing pipelines are responsible to (1) download, (2) reduce, and (3) convert the images found in the source data archives to a format suitable for use by the users in our service interfaces. All those processing steps run as individual jobs in a Kubernetes cluster of processing nodes. Each step of the pipeline is monitored, and a successful completion of the job allows the pipeline to proceed to the next step.

Planetary surface image observations are provided by the Planetary Data Systems[5] (PDS) nodes in data formats known as PDS Images, sometimes using more common formats such as JPEG-2000. These formats encapsulate raster data (pixel arrays) as well as metadata about the instrument and specific relevant information about the corresponding observation like the angle of the camera relative to the planet's surface or the position of the Sun relative to the spacecraft.

For Mars, the datasets we are providing come from two different spacecrafts, the Mars Reconnaissance Orbiter (MRO) and the Mars Express (MEX). The former, MRO, host different observation instruments which we are collecting data from, the Context Camera (CTX) and the High Resolution Imaging Science Experiment (HiRISE). Whereas MEX provides us with the data collected by the High Resolution Stereo Camera (HRSC). The level of data each dataset and the format the images are provided vary: MRO/CTX, for instance has to be fully reduced and map-projected before the user can use it for our proposes, while MRO/HiRISE and MEX/HRSC are provided as instrument calibrated.

[1] https://www.jacobs-university.de
[2] https://www.meeo.it
[3] https://adamplatform.eu
[4] https://explorer-space.adamplatform.eu
[5] https://pds.nasa.gov

As explained below, the primary job we carry out in NEANIAS is to homogenize all data sets regarding the image signal quality, map-projection and final image data format. Figure 1 represents the pipeline in finer grained steps as we use for our developments, “level-0” images is an alias for “source” data, data downloaded from the source data archives. Each image will through a series of processing steps as needed until it reaches “level-3”, previously called “high-level” data; GeoTIFF images.

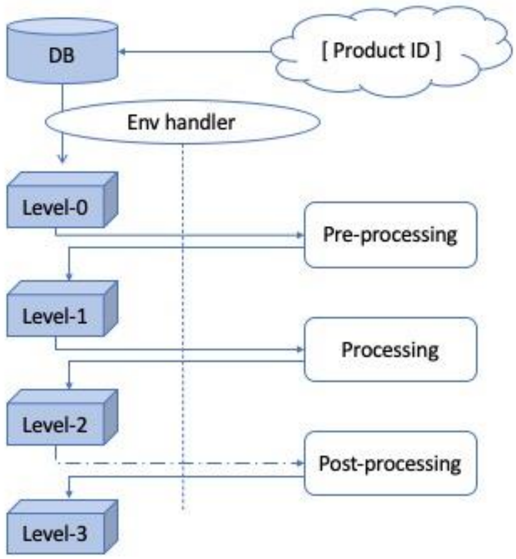

Figure 1: image data processing steps from download to data reduction to high-level data format by the ADAM platform.

Data download

The first step in our pipeline is to download the source data from each of the datasets. Image data products are downloaded individually from the PDS Geosciences Node[6]. The archive source location is retrieved by querying the Orbital Data Explorer (ODE) REST interface[7], queries are done on the basis of region on the planet’s surface – a bounding-box – to each typically many observations are identified. Since each image can have from a few dozens of Megabytes up to a couple of Gigabytes, the download itself of each individual image runs as a separate job in the queue system.

In *Figure 1*, this block is represented as “pre-processing”: as soon as an image is download an entry for it is inserted in the internal database (hosted by a MongoDB cluster instance) containing its location in the local storage as well as fundamental metadata like the geo-location as provided by ODE’s metadata. The (MongoDB) will be queried and updated accordingly at each step of the pipeline with the latest information for each image such as the location in the local storage (each processing step generates a new copy of each image, data is not overwritten).

[6] https://pds-geosciences.wustl.edu
[7] http://oderest.rsl.wustl.edu

Image processing

Once an image is successfully downloaded (and inserted in the data store), the *reduction* processing is launched, it is typically composed of many steps. In *Figure 1* we see this step as “processing”. In the case of MRO/CTX, the image reduction is composed by instrumental corrections applied to the image signal intensity. Whereas the other datasets are already provided in a calibrated state, living us with the task of projecting data to a global Mars Sinusoidal projection.

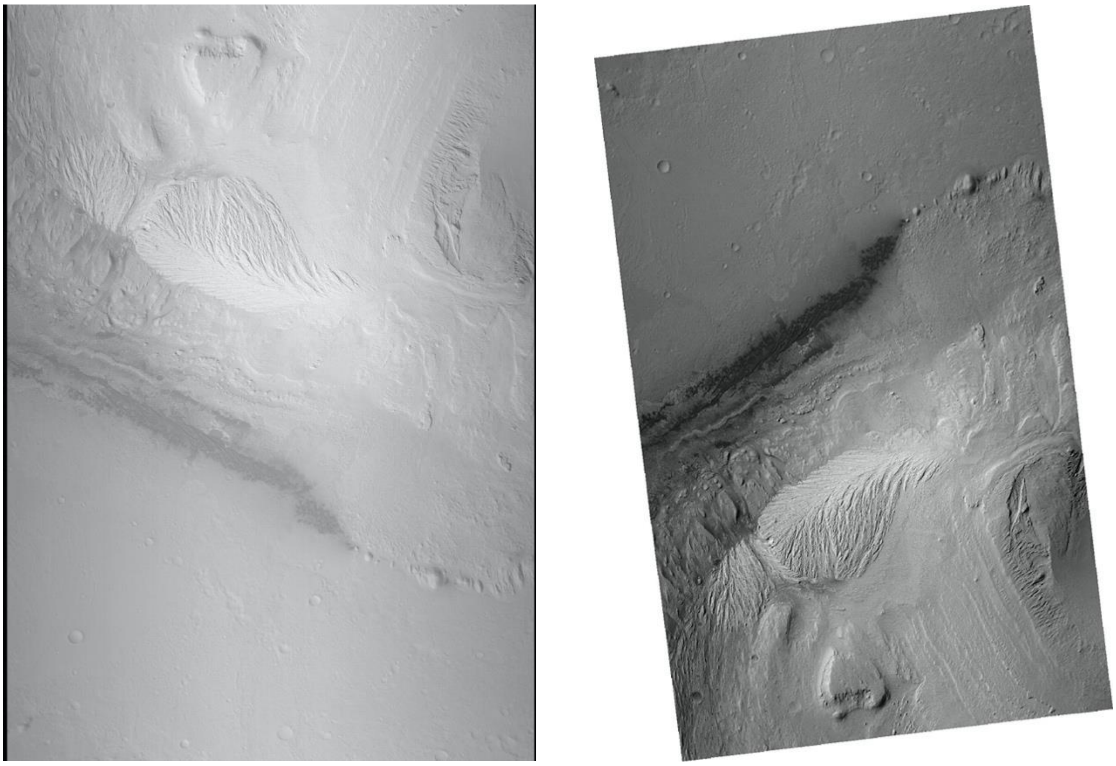

Figure 2: CTX B01_009861_1753_XI_04S222W. Left: the original PDS image. Right: map-projected and calibrated result.

Final product

Finally, we convert the image to a Cloud Optimized GeoTIFF (COG). This is the “post-processing” step in *Figure 1*. GeoTIFF (COG) allows for easy visualization and optimized transfer through web and a format that the user can download and operate at their will in any geo-aware software (e.g, QGIS). Those images are then displayed through ADAM-Space and available for download either through the graphical interface or through ADAM-API[8].

[8] https://docs.neanias.eu/projects/c2-dataexploration/